JOURNAL OF RESEARCH IN SCIENCE TEACHING VOL. 44, NO. 8, PP. 1108–1133 (2007)

Students' Cognitive Focus During a Chemistry Laboratory Exercise: Effects of a Computer-Simulated Prelab

T. Mikael Winberg, C. Anders R. Berg

Department of Chemistry, Umeå University, Umeå, Sweden

Received 27 June 2005; Accepted 3 April 2007

Abstract: To enhance the learning outcomes achieved by students, learners undertook a computer-simulated activity based on an acid–base titration prior to a university-level chemistry laboratory activity. Students were categorized with respect to their attitudes toward learning. During the laboratory exercise, questions that students asked their assistant teachers were used as indicators of cognitive focus. During the interviews, students' frequency and level of "spontaneous" use of chemical knowledge served as an indicator of knowledge usability. Results suggest that the simulation influenced students toward posing more theoretical questions during their laboratory work and, regardless of attitudes, exhibiting a more complex, correct use of chemistry knowledge in their interviews. A more relativistic student attitude toward learning was positively correlated with interview performance in both the control and treatment groups.

Keywords: chemistry; conceptual change; laboratory science

Many instructors believe that higher education laboratory exercises have the potential, not only to help students confirm, elaborate on, and put theoretical knowledge into a meaningful context, but also to help them learn scientific methodology and cultivate practical skills. Research, however, indicates that students often focus on manipulative details and other procedural issues rather than elaborating on the underlying theory and linking it to the exercise (Hofstein & Lunetta, 2004). The focus of students' cognitive activity during laboratory work is a key issue, as is their ability to discern significant information from "noise" and to process that information in a meaningful way. From cognitive load and working memory research, we know that the amount and functionality of a student's previous knowledge determines their ability to overview, filter, and process information (Baddeley, 1986; Johnstone, Hogg, & Ziane, 1993; Sweller, van Merrienboer, & Paas, 1998). Laboratory exercises, especially in higher education contexts, often involve training in several different manipulative skills as well as a high information flow, such as from manuals, instructors, output from the experimental equipment, and so forth. If students do not have prior experiences that help them to sort out significant information or reduce the cognitive effort

Contract grant sponsor: Natur och Kultur.
Correspondence to: T.M. Winberg; E-mail: Mikael.Winberg@educ.umu.se
DOI 10.1002/tea.20217
Published online 6 September 2007 in Wiley InterScience (www.interscience.wiley.com).

required to understand what is happening in the experiment, they tend to rely on working strategies that help them simply to cope with the situation; for example, focusing only on issues that are of immediate importance to obtain data for later analysis and reflective thought (Johnstone & Wham, 1982; Lundberg, 1998). Given the relatively large amount of time that is allocated to laboratory work in higher chemistry education, we argue that laboratory exercises must be used more efficiently to be fully justified from an economic as well as a pedagogical perspective. The question is: What we can do to help students maximize their learning during laboratory exercises?

Johnstone (1997) emphasizes the significance of pre-knowledge for effective laboratory-based learning. ‘‘First-time, unprepared learners are not in a position to process laboratory experiences with understanding’’ (p. 267), he wrote, claiming that what and how you learn is controlled by what you already know. He found that prelab exercises, involving a thorough revision of theory and reacquaintance with skills, significantly reduced the proportion of students' ‘‘thoughtless questions’’ during laboratory work and improved their performance on postlab assessments, suggesting that this might be a way to improve learning during laboratory work. Research on the nature of prelab activities used most frequently in schools and universities is not available. Furthermore, the focus of research on prelab activities found in research databases, such as ERIC, tells us more about trends in the research community than the kinds of prelab activities actually implemented in everyday teaching. From a cursory web search by Yahoo (USA) for ‘‘prelab,’’ resulting in 17,300 hits, it appears that prelab activities focusing on describing laboratory apparatus, procedures, and safety rules are most common, whereas theoretically oriented activities, focusing on reviewing central concepts and structuring theoretical knowledge of the domain, appear less frequently. Prelab instruction on the procedural aspects might reduce uncertainty about what is about to happen during an exercise. However, just reading through the laboratory manual or completing some calculations is unlikely to provide sufficient support for student construction of appropriate pre-knowledge, required to facilitate laboratory work (van Merrienboer, Kirschner, & Kester, 2003). Furthermore, we argue that prelab training is an opportunity for instructors to highlight what really matters during the exercise. Thus, a prelab exercise that gives students opportunities to develop experience in the domain and focuses on aspects considered to be of prime interest in the activity would not only facilitate processing of information but also influence students' cognitive focus during laboratory work.

Computer simulations, as well as paper-and-pencil exercises, have been proposed as a strategy for situating theoretical knowledge in the laboratory context and helping students actively review and elaborate on central concepts prior to their laboratory work (Johnstone, 1997). In this study, we evaluate the effects of a computer-simulated acid–base titration prelab exercise. First, the effects of the simulation on students' content knowledge are measured through interviews. Student attitudes toward learning have been shown to influence their learning (Hofer & Pintrich, 1997; Windschitl & Andre, 1998). Therefore, their impact on learning outcomes, in terms of content knowledge, is also examined. Second, differences in cognitive focus between simulation and non-simulation groups during the subsequent laboratory work are analyzed.

Literature Review

Cognitive load theory (CLT) is first reviewed briefly in what follows, presenting ideas on how students' previous knowledge and experience influence their ability to process information. Second, we discuss the different sources and the amount of cognitive load a laboratory exercise might impose on learners and, third, how computer simulations may help learners reduce cognitive load and hence allow more theoretical considerations in information-rich environments like laboratory exercises. Finally, the research on students' attitudes toward learning is reviewed to

illustrate the significance of considering this aspect of student characteristics when evaluating teaching interventions.

Cognitive Load

Laboratory exercises usually impose a high cognitive load on learners, impeding student performance during such work (Johnstone & Wham, 1982). According to the CLT (Sweller et al., 1998) and working memory research (Baddeley, 1986, 1992), working memory is where conscious ‘‘thinking’’ occurs, such as organizing, comparing, and elaborating on information. A central assumption is that working memory has only a limited capacity to process elements of information. Long-term memory, on the other hand, can store large quantities of information organized into schemata, which are treated as single information elements by working memory. For example, for a novice, the concept of buffer capacity could be made up of several separate schemata, comprising information such as the Henderson–Hasselbach equation, molecular structures, and pKa. For the expert, on the other hand, these elements make up a single structure where all elements are interrelated—a schema. Consequently, by constructing schemata, working memory limitations can be circumvented, allowing learners to allocate more capacity to learning (Kirschner, 2002). Functional schemata, that is, with a high degree of automation and complexity, also increase performance speed (Carlsson, Chandler, & Sweller, 2003) by making information more easily accessible (Sweller et al., 1998). The sources of cognitive load are generally classified according to whether they are related to learning or not, and whether they arise from the intrinsic properties of the task or the process of schema construction. Bannert (2002) summarized the three principal sources of cognitive load:

1. *Intrinsic cognitive load* (ICL) depends on the degree of interactivity among elements in the task and whether the learner is familiar with the domain or not. Familiarity with the domain, or having appropriate schemata, reduces ICL. Furthermore, extensive practice may lead to automation of schemata, allowing automatic, partly nonconscious, retrieval and application of procedures, which further reduces ICL and ‘‘speeds-up’’ problem-solving processes (Sweller et al., 1998). Consequently, experienced learners may engage in tasks with a large element of interactivity but still have working memory capacity left for schema construction (learning) or metacognitive activities.
2. *Extraneous cognitive load* (ECL) refers to cognitive load that is not related to the process of learning, or schema construction. In a laboratory setting, ECL could be caused by factors such as a poorly organized laboratory setting or an inappropriate instructional format. For example, provided that these are not the primary goals of the exercise, ECL could arise from a learner’s attempts to assemble the equipment, find a missing burette, or understand how a diagram should be read. Unclear, wordy, or otherwise redundant instruction is also a probable contributor to ECL.
3. *Germane cognitive load* (GCL) results from the learners’ effort to construct schemata, such as comparing and contrasting new information with existing knowledge, and is thus regarded as the cognitive load directly associated with learning.

These sources of cognitive load are assumed to be additive and non-discernible in practice. Thus, the load experienced, interfering with the limited capacity of working memory, is the sum of all sources (Sweller et al., 1998).

Cognitive Load in the Laboratory. For university chemistry education, planning and organization work and manipulative skills are important aspects of students’ learning to work productively in a laboratory. Although these aspects are not all manipulative, they will be referred

to as ‘‘practical’’ in this investigation. A second, equally important goal is to encourage students to expand on the theoretical knowledge treated in the exercise, such as paying attention to and reflecting upon their results or making theoretically grounded predictions. However, for mainly economic reasons, students are often expected to learn practical skills at the same time as building on theoretical aspects of the exercise, both tasks being potential sources of cognitive load. In later higher education science courses, at least, laboratory tasks are intended to be complex, reflecting the challenges a scientist could face in ‘‘real life.’’ Thus, in these situations, both ICL and GCL may be high. If the extraneous cognitive load is added—for example, one caused by malfunctioning or missing equipment—then the total cognitive load might be overwhelming. If working memory limits are exceeded during laboratory work, particularly given the fact that students must finish within specified time-frames, a cognitive focus upon practical activities in preference to theoretical considerations is to be expected (Johnstone & Wham, 1982).

To facilitate effective learning, including helping students to allocate some working memory capacity to the theoretical aspects of the laboratory exercise, we should consider how to lower cognitive load during laboratory work. We realize that ECL is sometimes hard to address within laboratory settings, so decreasing ICL might be a feasible alternative to improve the learning outcomes achieved regarding theoretical aspects of laboratory activity. Because intrinsic cognitive load partially depends on the ‘‘functionality’’ of learners’ schemata associated with the domain, proper student preparation (van Merrienboer et al., 2003) before laboratory activity is essential to minimize the sum of ICL and ECL, thus allowing for anticipated increases in GCL during laboratory work. The acquisition of expert-like schemata, with high complexity and some degree of automation, normally requires extensive experience. In higher science education settings, time is often too scarce to allow development of the necessary skills prior to laboratory exercises. Simulations have been proposed as a means to provide intense interactions with a topic, thus facilitating ‘‘tuning’’ of existing knowledge on a given topic into a more intuitive and readily accessible form in a relatively short time.

Learning Through Simulations

Often in simulations the central idea is to provide a complex, dynamic situation where students use their knowledge of the target concept to evaluate observations of the system, pose new questions, and design experiments to answer these questions (de Jong & van Joolingen, 1998). Swaak and de Jong (1996) claimed that students’ use of knowledge in complex situations initiates experiential learning processes that ‘‘tune’’ the knowledge and give it an intuitive quality, making it more rapidly accessible from memory and thus improving their ability to quickly evaluate complex situations. Swaak and de Jong (1996) described the intuitive quality of knowledge as ‘‘not easily verbalized’’ and difficult to assess using traditional methods that probe declarative knowledge (i.e., recall of facts and definitions). Instead, citing other studies, they proposed the latency time for answering questions and/or the ability to make quick predictions in complex situations as appropriate measures of students’ intuitive knowledge. This is consistent with the conclusions of Thomas and Hooper (1991), who asserted that tests designed to assess ability to implement knowledge in complex systems seem more appropriate for evaluating the effects of simulations than do tests focusing on declarative knowledge. Furthermore, they argued that simulations seem most effective when used in conjunction with formal training, indicating that learning gains could be attributable to modifying or tuning *preexisting* knowledge into a more efficient form.

It seems that what Thomas and Hooper (1991) and Swaak and de Jong (1996) described is analogous to CLT ideas of functional schemata. According to CLT, the ease with which

information is processed depends on appropriateness, complexity, and degree of automation of schemata (Carlsson et al., 2003). Because the acquisition of new explicit content knowledge is, apparently, unaffected to any great extent by simulations, we propose that the primary effect of simulations is the refinement of *existing* knowledge into functional schemata. Indeed, CLT predicts that tasks that have an open or explorative character, which are common in simulation exercises, allow for the development of schemata that lead to an increased ability to use knowledge in complex or new situations (Sweller, 1988). We propose that learning outcomes from simulations could be assessed by studying behaviors and abilities that are related to the degree of complexity and automation of students' schemata. Hence, observed differences in behavior or performance in high-complexity situations, such as laboratory work, or situations with a low degree of structure, such as semistructured interviews, would reflect differences in schema functionality. This idea is examined further in the Methods section.

Attitudes and Learning

Considerable research in science education has focused on learning outcomes for undifferentiated student cohorts. One characteristic that has been suggested as having an effect on learning outcomes is student attitude toward learning (Finster, 1989, 1991; Osborne, Simon, & Collins, 2003). In what follows we discuss attitude research based on Perry's model, including descriptions of attitudes considered relevant if learning outcomes for different students are the focus.

A considerable part of our theoretical thinking on development of student attitudes originates from William Perry's book, *Forms of Intellectual and Ethical Development During the College Years: A Scheme* (Perry, 1998). For an extensive overview of Perry's work and subsequent research, see Hofer and Pintrich (1997). Perry proposed a theory of intellectual and ethical development among college students, portraying a developmental process (not static personality traits) identifying nine stages or positions. These are usually condensed into four sequential categories (Fitch & Culver, 1984; Moore, 1994):

Dualism (Perry positions 1–2) is characterized by a dualistic right or wrong view of the world—authorities supposedly know the truth, which learners could acquire. The role of students is to learn the right answers.

Multiplicity (Perry positions 3–4) represents a modification of dualism, with the possible inclusion of "not yet known," in addition to right–wrong. The role of learners is to find knowledge and to think for one's self.

Contextual relativism (Perry positions 5–6) represents a major shift in perspective from a world with many exceptions to right or wrong, to the opposite view, that everything is relative and context-bound, with few right or wrong answers. Learners are active makers of meaning within a given context.

Commitment within relativism (Perry positions 7–9) mainly concerns an elaboration of identity and does not refer to cognitive change. Very few undergraduate students reach Perry positions 7–9 (Moore, 1994).

Perry's work was later modified (Finster, 1991; Fitch & Culver, 1984), and applied to science education (Mackenzie, Johnstone, & Brown, 2003). Finster (1991) adapted the Perry scheme to the context of chemistry education and presented examples of how a student's attitude position could affect their view of the roles of instructor, evaluation, and laboratory activities. A student holding a Perry-based contextual relativistic view would experience an open experiment or task differently, undergoing a greater challenge, than a student with a dualistic right–wrong view.

Other lines of research have focused on the question of attitude formation and change; however, attitude change is not addressed in this investigation, where our focus is on students' present attitudes and their relation to the outcomes of a prelab exercise.

Parallels Between CLT and Attitude Research

How tasks are perceived is addressed by both attitude research (Finster, 1991) and CLT; both approaches predict that the challenge is different for different students. How the task is perceived depends on attitude position; for example, ''Describe how buffer capacity depends on total concentration'' could be interpreted as ''I'm supposed to learn a formula given by my teacher to answer this,'' or ''I'm supposed to understand the relationship between concentration and buffer capacity with help from information in lectures and textbooks.'' Similarly, CLT predicts that ICL depends on the depth to which the student tries to solve the task. A student who perceives that the task is asking for more robust understanding will experience a higher ICL, whereas an interpretation of the task as ''I'm supposed to briefly understand'' would impose a lower ICL upon the student (given that schemata/knowledge have the same qualities within both students). Both of these complementary perspectives from CLT and attitude research can serve as background for a study of a prelab activity with open tasks (i.e., giving the students freedom to choose presentation level/perspective).

In this study we assess the ability of a computer-simulated prelab exercise to tune students' existing knowledge into more functional forms, and examine whether any positive effects on student knowledge are associated with a changed cognitive focus during subsequent practical laboratory work. Furthermore, we assess whether the simulation exercise is helpful, in terms of improving knowledge functionality, for students with *any* attitude toward learning, or if it is especially beneficial/detrimental for students with some specific attitude configuration. Given the explorative character of the simulation task, which is described later, and the characteristics of the different attitude positions associated with the Perry scheme, it was not apparent whether the simulation exercise would affect all students equally.

To our knowledge these issues have not been previously assessed in conjunction. Hence, besides being valuable for science educators, this study should be of interest to cognitive load researchers as well as those in the field of computer simulations, especially scientific discovery learning, and attitudes toward learning/epistemological beliefs.

Research Questions

The following three research questions are the basis of our study:

- What effect does exposure to the prelab simulation have on students' cognitive focus during laboratory work?
- How is the learner's knowledge functionality, in terms of tendencies and abilities to use their chemistry knowledge in an interview, influenced by exposure to the prelab simulation?
- How do students' attitudes toward learning affect their use of chemistry knowledge within an interview after the prelab simulation?

Methods

A course development project at our university, funded by the Swedish Council for Higher Education, revealed that students, although claiming they had mastered the procedures necessary

for solving school tasks, often lacked a global understanding of the domain. Also, during the laboratory exercise associated with the acid–base concept, students claimed that their primary goal was to gather useful data, and that thinking occurred afterwards while writing their lab report (Lundberg, 1998). To increase student engagement in theoretical aspects of laboratory work, a computer-simulated acid–base titration was introduced prior to the practical laboratory exercise. In this study, the effects of this prelab simulation are investigated.

Although the methods are discussed in greater detail later on, for clarity, we introduce them briefly at this point. The methods used to answer the research questions included:

1. Interviews, which provided information on how students use their chemistry knowledge in a relatively unstructured situation. This method was used to determine whether the simulation had any effects on the functionality of students' content knowledge.
2. Recording and classifying students' questions to the teacher during subsequent practical laboratory work. This method was used to determine whether the simulation had any effect on students' cognitive focus during laboratory work.
3. An attitude questionnaire, to determine students' attitudes, prior to the simulation exercise, toward learning chemistry.

Sample

The investigation consisted of two studies, both performed in a first-year, introductory chemistry course at Umeå University, Sweden. Admission to the course required previous studies in chemistry in upper secondary school. The course was 20 weeks long and students took no other courses during this period. The first study was completed in spring 2001 and the second in the following autumn.

Study 1. In the first study, the treatment group ($n = 78$) consisted of first-year students, following completion of their first chemistry course in 2001. This group comprised prospective biologists (57%), geologists (13%), chemical engineers (20%), and students who did not follow any specific program (10%).

The control group ($n = 97$) consisted of students who had been interviewed by the researchers during the same course in the previous year (2000). The prelab simulation was not part of the curriculum in this earlier course.

Cognitive focus during laboratory work was studied with full student groups ($n = 78$ and 97, respectively). For the interviews, students were selected based on their attitude questionnaire scores. A total of 16 students were selected to be interviewed: 5 from the treatment group who exhibited a more relativistic attitude toward learning (HiPos), and 3 from the control group who had more dualistic attitudes (LoPos) (Table 1).

Study 2. This second study was intended, in part, to check the reliability of the first study. Also, because both control and experiment groups were part of the same class, differences in previous experiences during the course were minimized. The group ($n = 58$) comprised prospective molecular biologists (31%), biologists (51%), geologists (7%), and students who did not follow any specific program (11%).

Students were divided into treatment ($n = 21$) and control ($n = 37$) groups. Distribution of students between the treatment and control groups was not controlled by the researchers. A total of 12 students were selected for interviews. Only two HiPos in the treatment group and three HiPos in the control group were identified (Table 1).

Table 1
Total number of students and students selected for interviews in Studies 1 and 2

	Total (n)	Selected for Interviews	
		HiPos	LoPos
Study 1			
Treatment group	78	5	3
Control group	97	5	3
Study 2			
Treatment group	21	2	3
Control group	37	3	4

Organization of the Study

As a part of the 20-week course, treatment and control groups in both studies underwent a 1-week instructional sequence consisting of lectures (10 hours), tutorials (2 hours), and problem-solving classes (8 hours) mainly covering the topics of buffers and solubility. Subsequently, a 1-week period designated for laboratory work was begun. The simulation was scheduled as one of five exercises during this laboratory week. While the treatment group completed the simulation, the control group performed another laboratory exercise, not concerned with pH and buffer capacity. Interviews were held immediately after the practical laboratory exercise that dealt with buffers. The intention was to avoid the learning effects of report writing, which, according to many students, is where the actual thinking occurs during lab exercises (Lundberg, Berg, Geladi, & Johansson, 2002). An overview of the experimental chronology and organization is presented in Figure 1.

The Prelab Simulation Exercise

For the simulation, students used WinSolGasWater (WinSGW), a computer program developed and used for complex equilibrium calculations in research at the Department of Chemistry, Umeå University, Sweden. To minimize operating problems, a new user interface was designed, limited to simulate only acid–base reactions in liquid and liquid/gas systems, called WinSGW-Light. Students were able to freely manipulate the starting volume, all concentrations,

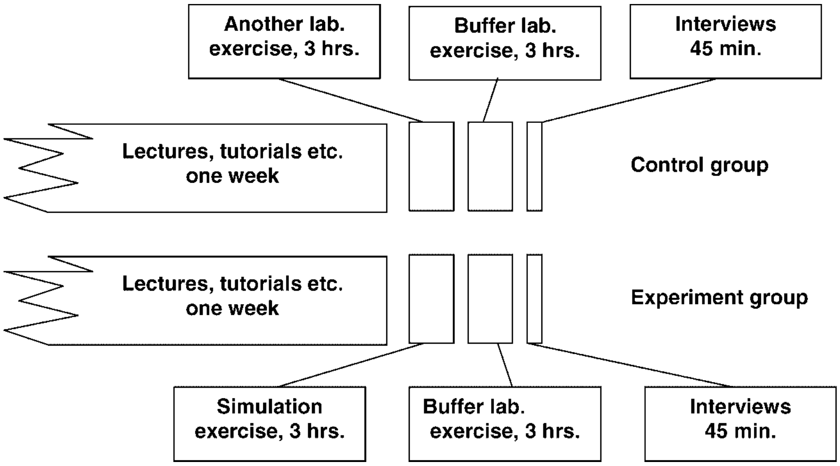

Figure 1. Overview of the study design.

pKa values, and the total acid in samples to be titrated with the strong base. It was also possible to begin with the corresponding base of a weak acid and titrate it with a strong acid. When prompted, the computer drew the titration curve as well as a diagram showing logarithms of the concentrations of species in the titrated solution within the interval pH 0 to pH 14 (Fig. 2). To compare experimental results with different settings, it was possible to display the information from several trials on the same titration diagram. For more information about the research version of WinSGW, or downloading of the demo version, see Lindgren and Karlsson (2001).

Before the simulation exercise started, all students received a 10-minute introduction, where the goal of focusing on qualitative aspects of content knowledge was emphasized. The "why's and how's" of data gathering were explained. The students worked in groups of two or three individuals. No lower time limit was set and the maximum time available was 3 hours. As a supportive framework for their work, students used written instructions containing 15 assignments, designed to initiate questioning and discussions rather than asking for explicit answers. Students' own hypotheses and experimental ideas were encouraged. No supportive capability was included in the program. Instead, the students were free to consult their notes and textbooks or ask fellow students in other groups or teachers who were present during the entire simulation session.

The theoretical concepts, ideas, and procedures addressed in the simulation exercise had been treated mainly quantitatively during the preceding 1-week instructional sequence; for instance, training on how to use acid ionization equilibriums and constants to construct pH diagrams that describe the speciation of the acids and bases involved over a pH range, or calculating pH in a

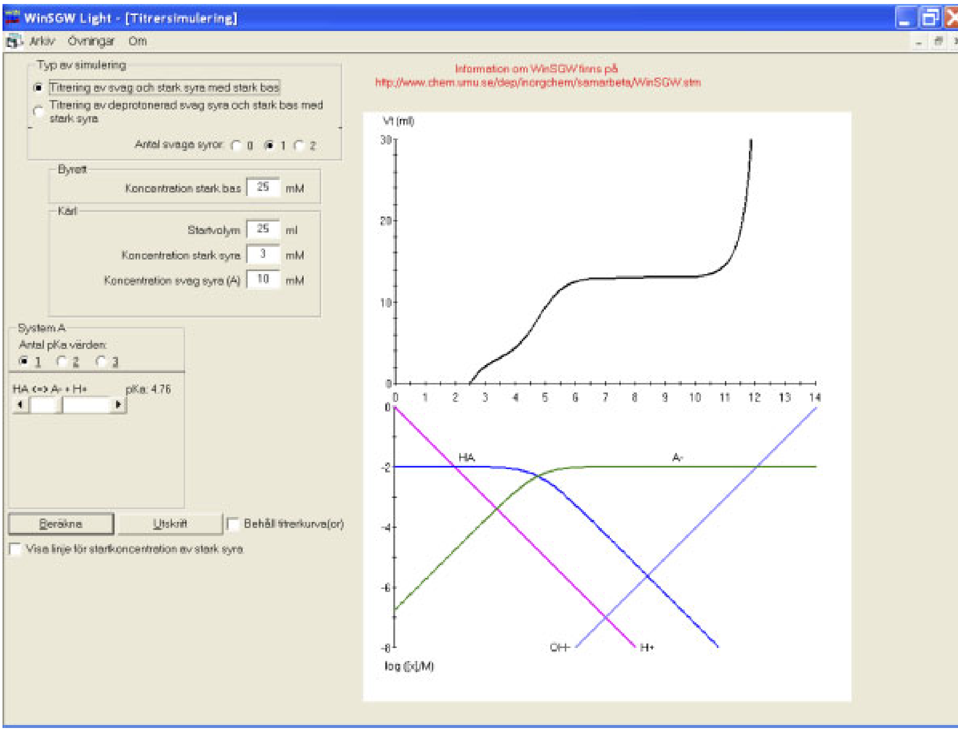

Figure 2. Example of diagrams associated with one of the 15 assignments, showing a titration curve (upper) and the corresponding pH diagram, also called a log diagram (lower).

given mixture of acids and bases. Because the goal was to improve students' qualitative understanding of the concepts and their interrelationships, assignments that could be completed purely by applying a formula were avoided. For example, the students were instructed to: "Discuss what characteristics a good titration curve would exhibit" (assignment 13); or "Discuss what information that could be inferred from the pH diagram" (assignment 1). In practice, no numerical values or quantitative relations were asked for. Hence, by using goal-free assignments, "means–ends" problem-solving strategies, known to impede learning (Sweller et al., 1998), were avoided. Known alternative concepts that had emerged in control-group interviews in Study 1 were used to create situations that would challenge these alternative conceptions, thus promoting learning discussions (Veerman, Andriessen, & Kanselaar, 2002) and joint efforts (Crook, 1995) to refine schemata. An example of this is assignment 14, where students were asked to describe what chemical reaction took place at just above pKa and at 2 pH units above pKa, respectively. The question was designed to examine the students' confusion between equivalence point and half-titration point. A gradual progression in complexity of the assignments was intended to reduce the risk of cognitive overload (Swaak, Van Joolingen, & de Jong, 1998; Sweller et al., 1998). The first assignments involved simple systems with a single pKa, whereas the last included several pKa's and the addition of strong acid. In addition to verbal interpretations, explanations, and evaluations, students had to write down their theories and conclusions. As a final assignment, they had to plan an investigation of a solution of their own choice for the second part of the laboratory exercise that followed on the day after the prelab exercise.

The Practical Laboratory Exercise

In the first part of the exercise, the task was to prepare a buffer with a predetermined pH. To accomplish this, students had to choose an appropriate weak acid and calculate the amount of strong base or acid to add to arrive at the desired pH. After the buffer was prepared, students titrated the solution with a strong acid and a strong base, respectively, and constructed a titration curve, thus obtaining a picture of the characteristics of the solution. In the second part, the students characterized a solution of their own choice to solve a problem or support/reject a hypothesis they had stated in advance. For example, one group compared the buffer capacity of fresh and sour milk, testing the hypothesis the buffer capacities would differ. Using the terminology of Domin (1999), the first part of the laboratory exercise was guided by instructions of an expository nature, whereas the second part was built on open inquiry.

Measurements

Cognitive focus during laboratory activities. Johnstone (1997) proposed that the frequency of thoughtless questions from students might serve as an indicator of student focus during experimental work. Although this method might provide information about whether laboratory work occupies students' minds, it does not reveal much about what aspects of the exercise do occupy their minds (i.e., their cognitive focus). Therefore, Berg, Bergendahl, Lundberg, and Tibell (2003) extended Johnstone's idea by observing whether the students' questions focused on theoretical or practical aspects of work and, furthermore, if their questions were reflective or spontaneous. In their study, Berg et al. showed that the character of the students' questions depended on whether expository or open-inquiry laboratory instructions were used, providing initial evidence for the predictive validity of the method. Later, Berg (2005b) showed that the students' questions to the teacher during laboratory work did indeed reflect what students reported as being the main focus of their mental efforts (i.e., cognitive focus). Student self-reports on the

focus and magnitude of their mental effort during the laboratory exercise correlated well with the group-level observations of student questions made by the assistant teachers. Based on these findings and similar research (Graesser & Olde, 2003; Kember, Wong, & Leung, 1999), Berg concluded that students' questions do hold evaluative qualities and that questions asked during laboratory work could be used as an indicator of their cognitive focus. The method does not require video cameras or microphones and the data collection is undertaken by the teacher while conducting normal supervisory work. Thus, the noninvasive character of the method makes it suitable for studying student behavior in a normal setting. For clarity we stress that the character of students' questions was used in the present study only to assess what the students were focusing on during the practical laboratory exercise—that is, what issues students invested mental effort in, and not the level of their chemistry knowledge.

Prior to the laboratory session, and after receiving instruction on classification illustrated by training questions, laboratory instructors practiced classifying authentic training questions recorded during earlier laboratory sessions. Training dealt with issues such as: "What should be classified as a theoretical question?" and "What is a thoughtful question?" Regarding classification of the practical/theoretical dimension, it was pointed out that a question with some theoretical content aimed at "just" being able to perform the practical task would be classified as practical. For example: a student referring to similarities in pKa's, asked: "Can I use methyl orange instead of bromophenol blue, which is specified in the instructions?" Regarding classification of the spontaneous/reflective dimension, it was stressed that sometimes it is difficult to immediately classify a question, but after discussion with the student the classification is easier. The classifications made by the instructors were then discussed and a common understanding of the dimensions was reached within the group. During laboratory work, instructors kept written records of the frequency and character of all questions in their laboratory group of 12–16 students, using the simple classification scheme shown in Appendix 1.

Knowledge functionality, as expressed in interviews. To assess whether the simulation had any effect on the functionality of student knowledge in the domain of acid–base and buffer reactions, semistructured interviews were conducted with the students immediately after they had completed their laboratory work, but before they wrote their laboratory report. Interviews lasted between 30 and 45 minutes, and were conducted by the authors.

Conversations are complex situations, there is an unspoken expectation to keep the conversation going, which, for example, involves attending to the other person's statements or questions, formulating adequate responses, predicting where the discussion is going, and providing new information or standpoints accordingly. van Bruggen, Kirschner, and Jochems (2002) provided support for the idea that interviews (i.e., synchronous communication) might indeed impose the necessary amount of cognitive load to discriminate between different levels of functionality of knowledge (schemata). From an affective perspective, having appropriate schemata would lead to a feeling of competence, which, besides the perceived degree of autonomy, is one of the cornerstones of intrinsic motivation (Ryan & Deci, 2000), most certainly influencing what interviewees elect to talk about when it comes to the use of chemistry knowledge in the interviews. We propose that, in a semistructured interview with a high degree of freedom, the student elects to talk about the subject in a way that feels comfortable, thus reflecting the ease with which information is processed—the appropriateness of the interviewee's schemata in the domain (Carlsson et al., 2003; Sweller et al., 1998). This aspect of knowledge has also been proposed as a measure of the outcome of simulations (Swaak & de Jong, 1996; Swaak et al., 1998; Thomas & Hooper, 1991). Thus, an increased spontaneous use of concepts and terms, as well as reference to interrelations between concepts, was expected to reflect the development of appropriate schemata during the computer-simulated prelab exercise.

Avoiding prestige bias, arising from the interviewees' own perceptions of how a statement would make them appear to the interviewer or themselves (Kvale, 1997), was of particular concern in this study. Making it clear to the students that they were being evaluated, or telling them what aspects we were interested in, could have made students focus on figuring out what we might want to hear or could have restricted their references to chemistry. Consequently, the risk that students would withhold important information about the level of their spontaneous use of chemistry knowledge would be high, threatening the validity of the interview data. Because the focus of the interview was to examine the accuracy and complexity of students' *spontaneous* use of chemistry knowledge, telling them the true purpose of the interview was obviously not an option. Instead, much effort was spent avoiding explicitly examining chemistry questions, but, at the same time, providing students with rich opportunities to use chemistry knowledge in their responses. Our impression was that, as long as we did not explicitly put pressure on them to explain specific chemistry concepts or phenomena, students were relaxed, answered openly, and used chemistry knowledge spontaneously. To gain rich information, the interviewer tried to keep all questions short, using silence to increase the pressure on the interviewee and letting the student talk as much as possible. For validity, interviewers asked follow-up questions and carefully probing questions whenever required, in order to clarify whether their interpretation was correct.

Between Studies 1 and 2, two new questions were introduced into the interview guidelines, intended to increase the proportion of "chemistry talk" by students. One real-life problem and one titration curve were introduced as stimuli for discussions. The interview guide is shown in Appendix 2.

Selection for interviews, attitude questionnaire. Because students' attitudes toward learning and knowledge were presumed to influence the outcome of the simulation, the selection of students for interviews was based on their responses to an attitude questionnaire. Students with low and high attitude positions were selected for interviews to maximize the contrast between groups. The attitude questionnaire was distributed at the beginning of each course. The attitude questionnaire was based on the work of William G. Perry (Perry, 1970) and subsequent applications in chemistry (Finster, 1991; Mackenzie et al., 2003) and constituted a further development of other extant questionnaires (Berg et al., 2003; Henderleiter & Pringle, 1999). The instrument consisted of 34 statements representing two viewpoints toward the attitude object (Reid, 2003). See Table 2 for illustrative statements.

Analysis of the responses to the attitude questionnaire was conducted using principal component analysis (PCA) (Eriksson, Johansson, Kettaneh-Wold, & Wold, 2001). PCA was used in the analysis because it is possible to position each student within the model described by the first, and only, significant principal component. A high score in the component indicates a relativistic attitude toward learning, here referred to as HiPos, whereas a low score corresponds to a more dualistic attitude, in this study termed LoPos. For the interviews, students with the lowest and highest attitude positions were selected, thus maximizing contrast. For a fuller description of the instrument, see Berg (2005a), wherein it was also shown that a positive change in attitude position, as measured by the instrument, is associated with motivated behavior and vice versa.

Interview analysis. The interviews were analyzed using the Qualitative Media Analyzer (QMA) (Skou, 2001), enabling direct association of coding marks with passages in the audio file, thus avoiding time-consuming transcripts with an inherent loss of situational information. All interviews from Study 1 were listened to several times. Passages judged to contain information about students' content knowledge were identified. These passages were sorted into some 20 working categories, in an attempt to capture the way that chemistry knowledge was used. As categories that were more comprehensive emerged, the total number of categories was reduced

Table 2
Illustrative items from the attitude questionnaire

		SA A N A SA	
1	Learning the material covered in the lectures should be enough to pass a course.	☐☐☐☐☐	Learning the material covered in the lectures is not enough to pass a course.
3	I think that lecturers should avoid including course material that is difficult for the students.	☐☐☐☐☐	I think that lecturers should include difficult course material to provide a challenge for the students.
19	I believe that I best learn the theory illustrated in the lab by planning and completing the experiment myself.	☐☐☐☐☐	I believe that I best learn the theory illustrated in the lab if there are explicit instructions showing how the experiment should be designed and completed.
30	It is important to include working with real samples e.g. ores or food during laboratory work even if it takes more time and is more complicated.	☐☐☐☐☐	To perform laboratory experiments with real samples is too time consuming and complicated to be worth the effort.

until only seven remained. These seven categories were used for analysis of student verbal discourse in both studies:

1. Misses the point, expressing misconceptions.
2. Describes the experimental procedure.
3. Mentions relevant concepts and/or ideas.
4. Comments on concrete aspects of phenomena.
5. Uses isolated concepts or ideas in a relevant way.
6. Uses two or more concepts or ideas, well integrated, in a relevant way.
7. Implements concepts or ideas outside the context defined by the course.

Some examples of how the scoring of interviews was achieved are described in what follows.

Category 1: "Misses the point, expressing misconceptions." There is clear confusion of concepts or the goals of the laboratory exercise. For example, the student claims, "at the equivalence point, the amounts of weak acid and its corresponding base are the same," thus confusing pKa with the equivalence point. Simple errors, such as saying "sodium carbonate" instead of "sodium hydrogen carbonate," are *not* coded in this category. An example of a student (S) "missing the point" is as follows (I = interviewer):

I: What would you have done if you had had another 3–4 hours available for practical work on your own experiment?
S: Eh...I think I would try to find more solutions to measure the pH in. I mean...it could be interesting to know what pH different solutions have.

This student has apparently perceived that the exercise was about measuring pH in different solutions. Because buffer capacity, or more precisely pKa, and concentrations of the weak acid system were central concepts in the exercise, the statement was coded in Category 1.

Category 2: "Describes the experimental procedure." This category sometimes included statements that were also coded in other categories. For example, a student might tell us what happened during the laboratory exercise and, during this, also use the concept buffer capacity in a relevant way, thus also scoring in Category 5 ("Uses isolated concepts or ideas in a relevant way").

Category 3: "Mentions relevant concepts and/or ideas." A student mentions a concept, for example, a buffer, but does not reveal any suggestion of actually understanding the meaning of the concept. For example, while offering an opinion on the laboratory exercise, one student said:

> It was valuable to get a bit more practical view on how buffers really work, as a complement to the textbook. . .to see how the pH diagram connects to the titration curve in a more practical fashion. . .because we saw that directly when we titrated our pH buffer.

Category 4: "Comments on concrete aspects of a phenomenon." One student noted, "If you put acid into a solution, the pH will drop," which is a correct statement, but it does not contain any relevant explanatory models.

Category 5: "Uses isolated concepts or ideas in a relevant way." For example, a student described the buffer capacity as the point were the system reacts with only small changes in pH when acid or base is added.

Category 6: "Uses two or more concepts or ideas, well integrated, in a relevant way." For example, a student associated the buffer capacity of a solution with the type and concentrations of the weak acid and its corresponding base, or with the total concentration of the system in a way that convinces the assessor that the student understands at least this aspect of the relationship between concepts. One student told the following story while describing what he would do if there had been more time for his own experiment:

> S: I would have tried to make a better buffer around pH 4. . .using other substances.
> I: Didn't your buffer work?
> S: Well, we used acetic acid and acetate. . .. It is a well-known fact were it buffers. . .I thought I should find something that buffered even better.
> I: Yes?
> S: I mean, we were supposed to make a buffer at pH 4. Acetic acid has its pKa at 4.76 so it doesn't buffer very well against acids—considering the symmetry, around pKa.

In a longer passage that follows, the student then refers to the pH diagram and the rapidly diminishing concentrations of acetate, convincing the assessor that he understands this relationship between pKa and buffer capacity.

Category 7: "Implements concepts or ideas outside the context defined by the course." An example, too long to relate here, is of a student who began reasoning about reactions in the coffee sample she had titrated, which smelled awful when made alkaline, but almost completely lost its smell when acidified. In her verbal reasoning, she referred to acid–base reactions of amino acids as well as the role of pH in the formation of volatile substances.

In general, a passage was defined as a continuous discussion of a specific topic, concept, or group of concepts in conjunction. Whenever a new concept was brought up, and not integrated with the ones discussed earlier in the passage, it was considered to be a new passage. The coding of passages was based on qualitative judgments of the meaning of students' statements, which often required the consideration of a wider context than the isolated passage. For example, a passage, although not conveying much evidence that the student understood the concepts discussed, could be classified as belonging to a relatively high category if the student had demonstrated

understanding of that concept in an earlier or subsequent passage. Another situation that required some caution was when the interviewer's questions were judged to have directly led the interviewee to respond in a certain way. In these cases, the response was not coded at all because it could not be regarded as being the result of the student's own choice of what topic and level of chemistry knowledge to use. The categories sometimes overlapped; in such cases, the student statement was coded with both categories. The authors coded all files independently; both authors have had extensive experience in teaching in the domain and had been involved in the initial development of the categories. After the individual coding of passages was completed, they met and discussed discrepancies until common viewpoints were agreed upon. Approximately 10% of the passages had to be discussed in this manner.

The inductive categories here share some similarities to the SOLO taxonomy developed by Biggs and Collis (1982), in that they indicate, to some extent, hierarchical aspects of observed content knowledge. A similar four-category system, comprising Categories 1, 5, and 6 from the present study, and one new category that described "the use of two or more concepts without relating them to each other," was also used by Winberg (2007) for coding contributions to threaded chemistry discussion groups on the internet. The study aimed at comparing the outcomes of three different methods of analysis: subjective characterizations and rankings made by university chemistry teachers, psychologists, and educational researchers; categorization of the observed use of chemistry knowledge (i.e., the system mentioned previously); and a 26-category system that focused on motivated behavior, interaction between discussion partners, metacognition, and interaction with the simulation, described in the present study. Data from the two latter methods were analyzed by PCA to obtain a ranking of discussions as well as distinguishing important categories. Results show that the characterizations of discussions correlated well between methods. In addition, the relative importance of categories for describing differences between discussions indicated a hierarchical relationship between the categories in the four-category system, with Categories 1 and 6 from the present study being the most important and being negatively correlated. These results provide support, not only for the validity of the coding system used in the present study, but also for the appropriateness of analyzing qualitative data statistically.

When considering the distribution of counts for the categories within interviews, a pattern of students' conversation emerges, which tells us something about how these students used their knowledge in the interviews. We argue that this pattern of students' content knowledge use may reflect the degree of usability of that knowledge, that is, the functionality of schemata. The relative occurrences and number of passages of categories found in interviews are presented in Table 4. In the Results section the patterns between different groups of students are compared statistically using chi-square tests based on counts.

Results

Engagement in the Simulation Exercise

The students, despite our stipulation of no minimum time on the task, used almost all the available time. Students also simulated approximately three times more scenarios than the 15 specified in the instructions (Table 3). Log files (not shown), as well as our own observations, indicate that most of these additional simulations dealt with their own hypotheses, clarifying and/or testing assumptions, and trying extreme values (very high or low concentrations, etc).

Our general impression was that student discussions were lively and focused. The researchers present during the simulation overheard only chemistry-related conversations among students. Student questions concerning practical issues, like handling the software or the computer, were

Table 3
Mean number of calculations and time on task (with standard deviation)

	Total Calculations		Time on Task (minutes)	
	Mean	*SD*	Mean	*SD*
Study 1	44	18	140	42
Study 2	53	24	161	56

few in number and were limited to the first 20 minutes of work. After that, questions for teachers were almost exclusively aimed at understanding the chemistry connected with the simulation exercises.

Cognitive Focus During Laboratory Work

Study 1. During the laboratory exercise, the treatment group asked a higher proportion of reflective questions regarding underlying theory than did the control group (Fig. 3). Differences in the distributions of questions according to category were not significant at the 5% level, using a chi-square test (chi-square = 7.526, $df = 3$, $p = 0.057$).

The total numbers of student questions were 173 and 154 in the control and treatment groups, respectively, producing an average of 1.8 and 2.0 questions per student.

Study 2. Differences between treatment and control groups with respect to the distributions of questions according to category were significant (chi-square = 30.496, $df = 3$, $p = 0.000$). According to chi-square values, the major differences resulted from the higher counts of theoretically oriented questions and lower counts of spontaneous practical questions in the treatment group (total chi-square = 23.016). As in Study 1, the proportion of reflective questions was higher in the treatment group than in the control group (Fig. 4). There was also a higher proportion of spontaneous theoretical questions in the treatment group. The total number of questions was 167 for the control group (n = 57) and 44 for the treatment group (n = 21), producing an average of 2.9 and 2.1 questions per student, respectively.

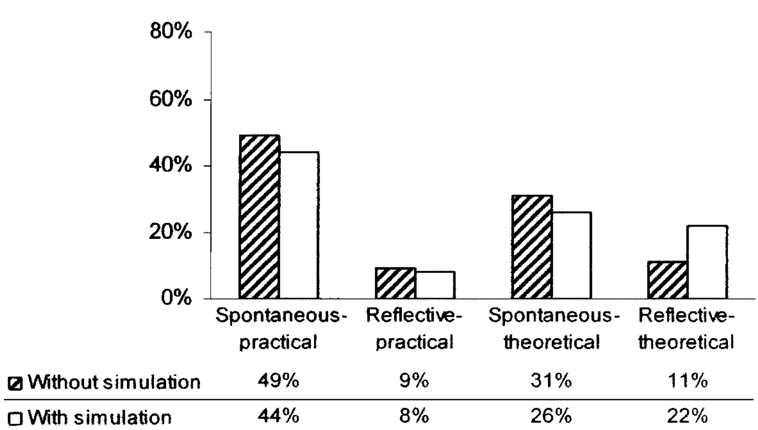

Figure 3. Relative distributions of question types during laboratory work in Study 1. Comparisons are between treatment (with simulation) and control (without simulation) groups.

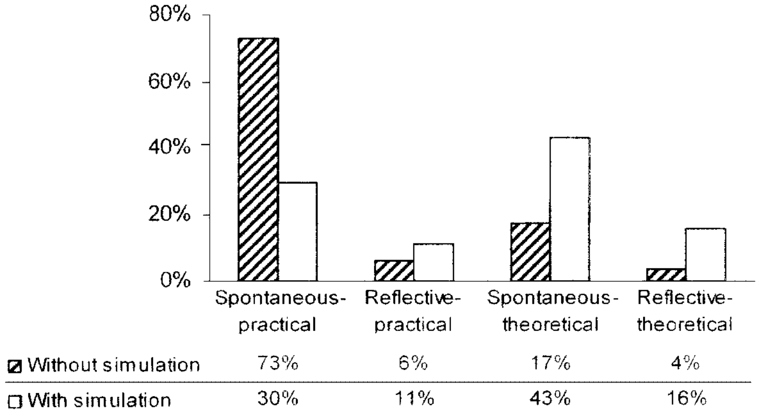

Figure 4. Relative distributions of question types during laboratory work in Study 2. Comparisons are between treatment and control groups.

Knowledge-Usability Interviews

Study 1. A total of 16 students were interviewed, 8 each from the treatment and control groups. The average number of coded passages per student was 13.5 in the treatment group and 9.4 in the control group.

Distributions of passages on the basis of category were significantly different between the treatment and control groups (Table 5, groups A and B). Treatment-group students exhibited a lower proportion of misconceptions (none) and a higher proportion of statements in the pooled Categories 6 + 7 (“Uses two or more concepts or ideas, well integrated, in a relevant way” and “Implements concepts or ideas outside the context defined by the course”) than did the control group (Table 4). The absence of passages in Category 1 (“Misses the point, expressing misconceptions”) in the treatment group strongly contributes to the significance of the chi-square test result. When studying the results from the interviews regarding students’ attitudes toward knowledge and learning, similar patterns emerge. Both HiPos (Table 5, subgroups C and D) and

Table 4
Distributions of coded passages for each category[a]

	Selection	Subgroup	Simulation	Category 1	2	3	4	5	6	7	Total Passages
Study 1	All	A	No	8	23	37	19	11	1	1	75
		B	Yes	0	17	46	14	12	10	1	108
	HiPos	C	No	4	21	36	21	14	2	2	56
		D	Yes	0	11	47	15	11	15	1	73
	LoPos	E	No	21	26	42	11	0	0	0	19
		F	Yes	0	29	46	11	14	0	0	35
Study 2	All	G	No	8	15	21	15	26	12	3	154
		H	Yes	17	12	22	9	32	7	0	81
	HiPos	I	No	6	18	20	14	24	14	4	71
		J	Yes	11	6	19	8	39	17	0	36
	LoPos	K	No	11	12	22	16	28	10	2	83
		L	Yes	22	18	24	9	27	0	0	45

[a]Data expressed as percent of the total number of coded passages in the groups in Studies 1 and 2.

Table 5
Comparisons of distributions of passages between groups in Studies 1 and 2 (see Table 4 for definitions of A–L)

	Comparison	Chi-square	*df*	*p*	Difference[b]
Study 1	A and B	15.143	5[a]	0.010	Yes
	C and D	10.393	4[a]	0.034	Yes
	E and F	10.150	4[a]	0.038	Yes
	C and E	10.604	4[a]	0.031	Yes
	D and F	10.020	4[a]	0.040	Yes
Study 2	G and H	9.856	6	0.131	No
	I and J	7.851	6	0.249	No
	K and L	9.766	5[a]	0.082	No
	I and K	3.615	6	0.729	No
	J and L	12.511	5[a]	0.028	Yes

[a]Passages in categories pooled to avoid expected counts below 1. Categories with expected counts of less than 5 were accepted when average expected counts were higher than 6 at the 5% level of significance (Zar, 1999).
[b]Significant at the 95% level.

LoPos (Table 5, subgroups E and F) achieved significantly more, on average, in the treatment group than they did in the control group. Chi-square statistics (not shown) indicate that Categories $1+2$ and $6+7$ are important for describing differences between treatment and control groups. Also, HiPos students generally had better-than-expected achievement in the interviews, when compared with LoPos students. Differences between HiPos and LoPos responses were significant within both treatment (Table 5, subgroups D and F) and control groups (Table 5, subgroups C and E).

Study 2. A total of 12 students were interviewed, 5 in the treatment and 7 in the control groups, respectively. Average numbers of coded passages per student were 16.2 in the treatment group and 22 in the control group.

Between the treatment and control groups, there were no significant differences in distribution of passages on the basis of category (Table 5, groups G and H). Also, there were no significant differences in distribution of passages for HiPos (Table 5, subgroups I and J) or LoPos (Table 5, subgroups K and L) when comparing treatment and control groups. In the treatment group, there were significant differences between HiPos and LoPos responses during the interviews (Table 5, groups J and L). In this comparison, the main differences, indicated by chi-square contributions, were in Category 6 (‘‘Uses two or more concepts or ideas, well integrated, in a relevant way’’), where HiPos students were responsible for more passages than expected, whereas LoPos students were responsible for fewer. In the control group, no significant differences were found between HiPos and LoPos (Table 5, subgroups I and K).

Analysis of Group Composition

Student achievement in the interviews in Study 2 indicated that, in addition to attitude toward learning, some other factor might have influenced the learning outcomes of the simulation. Therefore, a retrospective check of examination results for treatment and control groups was conducted (Table 6). In essence, we conclude that *interviewed* students in the treatment group in Study 2 performed significantly less well than the control group, whereas there were no differences in Study 1. For the treatment and control groups *where questions within the lab were recorded*, no differences in examination results were found in Studies 1 or 2. That is, with respect to examination results, the interviewed students in the treatment group in Study 2 were not

Table 6
Mean examination results for treatment and control groups in Studies 1 and 2

	Study 1		Study 2	
	Mean	*SD*	Mean	*SD*
Interviewed				
Treatment	20.5	6.2	14.8	3.4
Control	20.4	10.1	23.2	4.9
Questions in the lab				
Treatment	24.4	7.4	20.5	7.3
Control	23.4	8.1	23.4	7.2
All students	23.8	7.8	22.6	7.8

representative of the whole treatment lab group in this study. In Study 1, the interviewed students in the treatment and control groups were representative of their respective lab groups.

As an ordinary part of the course in Study 2, six examinations were undertaken at different times throughout the course; Study 1 had only one examination, which was done at the end of the course. A correlation analysis was performed on the results of all students in the six examinations in Study 2. A high degree of correlation could be found between all examinations ($p < 0.0005$ for all comparisons, correlation coefficient ranging from 0.562 to 0.728). Thus, differences in students' end-of-course examination results seem to have been consistent throughout the whole course. We propose that examination results in both studies were representative of any differences in previous knowledge between the control and treatment groups at the time of the simulation.

There was some skewing of the test results between interviewed students in the treatment and control groups in Study 2. With respect to students' average performance in the six examinations, the control group ($n = 7$) performed significantly better (two-tailed *t*-test, $p = 0.008$) than the treatment group ($n = 5$). In Study 1, there was no significant difference (two-tailed *t*-test, $p = 0.98$) in examination performance between the treatment ($n = 8$) and control groups ($n = 8$).

With respect to their examination results, the interviewed students in the treatment and control groups in Study 1 were representative of their respective whole lab groups ($t_{0.05\ (2)}, = -1.57, p = 0.168$ and $t_{0.05\ (2)}, = -0.64, p = 0.555$, respectively) examined for questions during the lab. In Study 2, the interviewed students in the control group were representative of the lab group on all tests, whereas the treatment group deviated significantly from the lab group, scoring significantly lower in four of six tests (Table 7).

In what follows is a summary of the information found in response to the research questions:

What effect does exposure to the prelab simulation have on students' cognitive focus during laboratory work?

Table 7
Results of two-sample t-test comparing examination results between interviewed and all lab-group students in the treatment group in Study 2

Test	*t*	*p*
1	3.00	0.029
2	2.60	0.060
3	3.00	0.040
4	3.21	0.049
5	3.02	0.094
6	3.41	0.011

In both studies, the treatment group exhibited a higher frequency of reflective questions regarding underlying chemistry theory than did the control group (Figs. 3 and 4). In Study 1, the difference was a doubling of the proportion of reflective theoretical questions between treatment and control groups. In Study 2, the difference was larger: proportionally, the treatment group asked four times more reflective theoretical questions than the control group. The proportion of theoretical questions (spontaneous and reflective) was almost three times higher in the treatment than in the control group. Results imply that treatment-group students focus more on the theoretical aspects of what they are doing. One way to explain this is that students within the treatment group, by completing the simulation, possess more complete schemata and hence have cognitive capacity available for thinking about what they are doing.

How are students' tendencies and abilities to use their chemistry knowledge in an interview influenced by exposure to the prelab simulation?

Study 1. The treatment group exhibited more frequent use of explicit chemistry knowledge, and were responsible for a higher frequency of passages where they, correctly, connected different aspects of theory. This can be seen in the higher frequency of Category 5 and Category 6 passages (Table 4). In addition, students in the treatment group were responsible for no Category 1 passages (misses the point, expressing misconceptions). Results imply that students who completed the simulation used chemical concepts, both isolated and integrated, more readily when talking about the laboratory exercise. This is in accord with our results from studies of cognitive focus during laboratory work, where theoretical and reflective questions were more commonly asked by the treatment group (Fig. 3).

Study 2. In interviews with the 12 students in Study 2, no significant effects regarding student use of chemical terminology and concepts were found (Table 5).

How do students' attitudes toward learning affect their use of chemistry knowledge within an interview situation after the prelab simulation?

Study 1. From what has been described, the simulation had a positive effect on the usability of theoretical chemistry knowledge. This positive effect was true both for HiPos and LoPos students. For HiPos, the main difference is that the treatment group was able to use and connect central chemistry concepts more extensively than the control group (Tables 4 and 5, subgroups C and D). For LoPos, the treatment group also scored higher than the control group (Tables 4 and 5, subgroups E and F). The differences were that the treatment group used isolated chemistry concepts in a relevant way and did not reveal any faulty chemistry knowledge. The control group was more focused on practical aspects of their laboratory work, mentioning chemistry concepts or ideas only when describing procedures and phenomena. Both HiPos and LoPos performed, on average, one category better in the treatment group than in the control group.

Within the treatment group, HiPos performed significantly better than LoPos in interviews. HiPos students generally reached one category higher than LoPos students and they revealed no faulty knowledge (Category 1). Essentially the same pattern was observed within the control group.

Study 2. Due to an imbalance between treatment and control groups with respect to school achievement (Table 6), and probably also with regard to content knowledge at the time of the interviews, we believe it is inappropriate to discuss these data herein.

Discussion

Simulation Caused Schema Development

Results for the first research question—the effect of simulation on cognitive focus during laboratory work—shows that the simulation resulted in a more theoretical focus during laboratory

work. This was indicated by more questions of a theoretical nature and also more questions that were reflective and theoretical. In the first study, it could be argued that this effect was caused by the students in the control and treatment groups having different experiences during the course; that is, these groups came from courses presented on two different occasions. For this reason, the second study was performed on students originating from the same course, specifically those having similar experiences. The persistence, and even reinforcement, of the trend toward theoretical, especially theoretical and reflective, questions indicates that the simulation had an effect on student cognitive focus during laboratory work. We propose two possible major causes for this effect. First, students may have associated the clear theoretical focus in the simulation with the subsequent laboratory exercise and thus perceived theoretical considerations to be more important than practicalities during the laboratory work. Second, as could be predicted from CLT, more functional schemata would reduce ICL, thus freeing working memory capacity and making theoretically oriented considerations possible in a laboratory setting with a high cognitive load. The first explanation, however, assumes that students have working memory capacity left for these theoretical considerations, an assumption that has been challenged by Johnstone (1997). Furthermore, the results associated with the second research question, at least in Study 1, indicate that the treatment group students did, indeed, develop functional schemata. This development was manifested by the more frequent and integrated use of chemistry concepts in interviews. As argued in the Methods section, interviews seem to impose sufficient cognitive load to reveal any differences in functionality of student schemata. Schemata comprising adequate, well-structured information facilitate information retrieval and speed-up performance (Carlsson et al., 2003; Sweller et al., 1998). This could explain the treatment group's better ability to discuss chemistry in the interviews. Thus, we argue that the change to a more theoretical focus during the laboratory task could be explained by the development of more functional schemata during the prelab simulation, and also by the simulation, sending a signal that theoretical considerations are important.

All Students Benefit From the Simulation

As described in the Results section, both students with more relativistic attitude positions (HiPos) and those with more dualistic attitudes (LoPos) benefited from the prelab simulation. The difference between groups was that, in interviews, HiPos reached higher in our categorization than LoPos (Table 5). Two interrelated plausible explanations can be advanced for this difference between groups. First, during the simulation, HiPos students could have focused more on exploring relationships among concepts than LoPos students, who were more focused on exploring isolated concepts. It is perhaps not surprising that students with a relativistic attitude are more likely to profit from an open learning situation with a high degree of freedom and that they consequently develop a more complex understanding than do students possessing a more dualistic attitude toward learning. Support for this idea was reported by Windschitl and Andre (1998), who found that students with relativistic-type beliefs learned more from simulations with a high degree of freedom than those with dualistic-type beliefs. The second possible explanation is that HiPos could, prior to simulation, have developed a more coherent understanding of the concepts and hence started at a different level. An indication that this might be the case can be seen in the control group, where HiPos showed a fuller understanding than LoPos in the absence of the simulation. Taken together, we argue that students' attitude positions affect their learning outcomes, but it is important to remember that both HiPos and LoPos display learning gains, although at different levels.

Pre-Knowledge is Important

As described in the Results section, Study 2 did not fully duplicate the findings in Study 1. Both studies show a more theory-oriented focus among students during laboratory work (Figs. 3 and 4). In interviews, on the other hand, no significant differences were found between experimental and control groups in Study 2, unlike Study 1, where the experimental group showed higher-level performance. In a retrospective check of examination results for Study 2 students, it was found that interviewed treatment-group students performed, on average, significantly less well in course exams than the control group students. For students in Study 1 no such difference in performance was found (Table 6). The examination was not specifically about acid–base chemistry, but covered the entire semester of chemistry studies and, hence, gives a general measure of their chemistry performance. It could be speculated that, due to weaker formal pre-knowledge—that is, missing the "bits of information" to incorporate into schemata—the treatment group could not make full use of the simulation, adversely influencing their performance in subsequent interviews. This result supports the idea that the primary learning gain from the simulation is refinement of schemata using previously learned "bits of information," or "tuning" of students' existing knowledge, rather than gaining new factual knowledge of the domain. The results are not surprising because the simulation exercise was aimed at clarifying and structuring existing knowledge. Apparently, students' level of pre-knowledge is an important factor, determining the outcome of the simulation. This is in accordance with the claim that the level of pre-knowledge affects student performance during simulation exercises (de Jong & van Joolingen, 1998).

It is debatable whether the interviewed students were representative of all students in the treatment and control groups. However, the high correlation between the examinations in Study 2 indicates that the differences in chemistry knowledge in the first examinations persisted throughout the whole course. If this finding is also applicable to Study 1, then the end-of-course examination in this study might have been representative of any differences between students' knowledge prior to the simulation. Assuming this is the case, the results show that the interviewed students in the treatment and control groups in Study 1 were representative of all the students in the treatment and control groups, among whom a shift in cognitive focus during laboratory work was observed. In Study 2, the interviewed students in the treatment group were not representative of all students that had undertaken the simulation exercise. Because a change in cognitive focus during laboratory work was observed in this larger group, we argue that there had been an improvement in student schemata, even though this could not be shown in the interviews because of the less knowledgeable students in this group.

Implications for Teaching

As previously mentioned, laboratory tasks are sometimes deliberately complex. The assumption is that such tasks support integration of theoretical knowledge with different skills and attitudes, ultimately enabling transfer to real-life tasks (van Merrienboer et al., 2003). Complex tasks impose potentially high cognitive loads, threatening to thwart any learning during the exercise. Thus, it is important to identify efficient strategies for managing cognitive load during complex learning. As general advice, van Merrienboer et al. (2003) proposed that supportive information necessary for problem solving and reasoning, due to its frequent high intrinsic complexity, should be provided before the task. On the other hand, schemata concerning procedures, such as how to assemble titration equipment, are mainly constructed and strengthened

by practice. Consequently, advance elaboration of manipulative procedures has little value, and should be provided "just in time."

In this study, the prelab simulation aimed at refining students' schemata with respect to the theory central to the laboratory exercise, and thus reducing the intrinsic cognitive load experienced by students (Sweller et al., 1998). We argue that the observed shift in students' cognitive focus toward reflection about theoretical issues during laboratory work is partially attributable to the freeing-up of some of their working memory capacity. If the only effect of the prelab simulation was to lower cognitive load, more reflective student questions regarding practical issues would also have been expected. As this was not the case, we propose that the focus of the prelab also provides a signal about what is important in the subsequent practical exercise. Thus, for teachers designing prelab exercises, it is not only important to consider how the prelab could facilitate student lab work but also how the prelab might affect students' perceived aims of the laboratory activity.

In conclusion, prelab exercises intended to help students integrate their theoretical content knowledge into schemata can help to make room for reflection but may also provide students with a sense of direction about what to attend to during their laboratory work.

The authors thank Professor Henry Heikkinen, University of Colorado, for extensive comments on previous drafts, and Assistant Professor Bruno Lundberg for initial discussions about the project and construction of the tasks used in the simulation exercise. Dr. Johan Lindgren constructed the WinSGW-Light user interface. We also thank Assistant Professor Leif Hedman and Professor Johan Lithner for providing valuable support during the writing process.

Appendix 1

Scheme used for recording the frequency and character of student questions during laboratory work (Fig. 5).

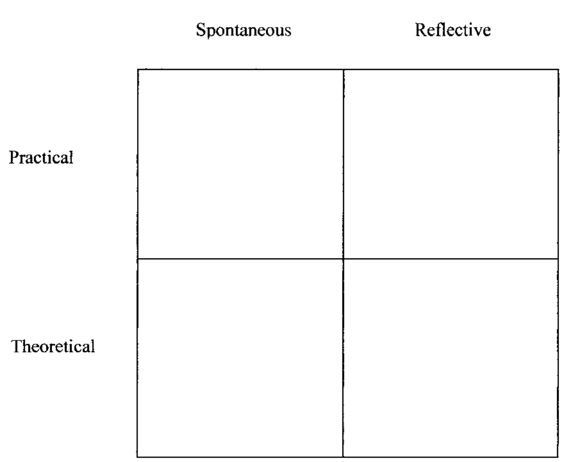

Figure 5. Scheme for recording frequency and nature of student questions during laboratory work.

Appendix 2

Interview Guide 1

1. *"Warming-up" phase*. Welcoming the student and briefly describing that the interview is about the laboratory exercise the student has just conducted. Brief information on issues regarding personal integrity and conditions for participation were given.
2. *Interview phase*. These questions were covered during the interview, sometimes in a different order or with different wording due to the development of the interview (note: the two questions in italics were used only in Study 2):

- What is the first thing that comes to mind regarding the laboratory exercise that you have just completed?
- Tell me what you did during this laboratory exercise.
- What is your opinion of the exercise?
- How did you prepare yourself for the exercise?
- What do you think you have learned from the exercise?
  - Was there anything that facilitated learning?
  - Was there anything that complicated learning?
- What would you have done if you had had an extra half-day to continue the practical laboratory work?
- *Suppose that you took a water sample from the creek that flows through campus.*
  - *Do you have any hypotheses about the contents of the sample?*
  - *How would you design tests required to characterize the sample with respect to buffer capacity, etc?*
- *What does this diagram tell you?(Interviewer shows several titration curves with corresponding pH diagrams of increasing complexity.)*
- What do you think was the reason for including this laboratory exercise in the course?
- Do you think you have learned anything essential thus far during this course?

3. *Verifying phase*. Did the interviewer correctly interpret what the student said?
4. *Debriefing phase*. Thanking the student for the interview and saying that we will listen to the interview and evaluate what was said, but no personally identifiable comments or statements will be reported.

References

Baddeley, A. (1986). Working memory. Oxford, UK: Oxford University Press.

Baddeley, A. (1992). Working memory. Science, 255, 556–559.

Bannert, M. (2002). Managing cognitive load—recent trends in cognitive load theory. Learning and Instruction, 12, 139–146.

Berg, A. (2005a). Learning chemistry at the university level: Student attitudes, motivation, and design of the learning environment. Unpublished doctoral dissertation, Umeå University, Umeå.

Berg, C.A.R. (2005b). Factors related to attitude change toward learning chemistry among university students. Chemistry Education Research and Practice, 6, 1–18.

Berg, C.A.R., Bergendahl, V.C.B., Lundberg, B.K.S., & Tibell, L.A.E. (2003). Benefiting from an open-ended experiment? A comparison of attitudes to, and outcomes of, an expository versus an open-inquiry version of the same experiment. International Journal of Science Education, 25, 351–372.

Biggs, J.B., & Collis, K.F. (1982). Evaluating the quality of learning: The SOLO taxonomy (structure of the observed learning outcome). New York: Academic Press.

Carlsson, R., Chandler, P., & Sweller, J. (2003). Learning and understanding science instructional material. Journal of Educational Psychology, 95, 629–640.

Crook, C. (1995). On resourcing a concern for collaboration within peer interactions. Cognition & Instruction, 13, 541–546.

de Jong, T., & van Joolingen, W.R. (1998). Scientific discovery learning with computer simulations of conceptual domains. Review of Educational Research, 68, 179–201.

Domin, D.S. (1999). A review of laboratory instruction styles. Journal of Chemical Education, 76, 543–547.

Eriksson, L., Johansson, E., Kettaneh-Wold, N., & Wold, S. (2001). Multi- and megavariate data analysis: Principles and applications. Umeå: Umetrics.

Finster, D.C. (1989). Developmental instruction part 1: Perry's model of intellectual development. Journal of Chemical Education, 66, 659–661.

Finster, D.C. (1991). Developmental instruction part 2: Application of the Perry model to general chemistry. Journal of Chemical Education, 68, 752–756.

Fitch, P., & Culver, R.S. (1984). Educational activities to stimulate intellectual development in Perry's scheme. Paper presented at the ASEE.

Graesser, A.C., & Olde, B.A. (2003). How does one know whether a person understands a device? The quality of the questions a person asks when the device breaks down. Journal of Educational Psychology, 95, 524–536.

Henderleiter, J., & Pringle, D.L. (1999). Effects of context-based laboratory experiments on attitudes of analytical chemistry students. Journal of Chemical Education, 76, 100–106.

Hofer, B.K., & Pintrich, P.R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. Review of Educational Research, 67, 88–140.

Hofstein, A., & Lunetta, V.N. (2004). The laboratory in science education: Foundations for the twenty-first century. Science Education, 88, 28–54.

Johnstone, A.H. (1997). Chemistry teaching: Science or alchemy? Journal of Chemical Education, 74, 262–268.

Johnstone, A.H., Hogg, W.R., & Ziane, M. (1993). A working memory model applied to physics problem solving. International Journal of Science Education, 15, 663–672.

Johnstone, A.H., & Wham, A.J.B. (1982). The demands of practical work. Education in Chemistry, 19, 71–73.

Kember, D., Wong, A., & Leung, D.Y.P. (1999). Reconsidering the dimensions of approaches to learning. British Journal of Educational Psychology, 69, 323–343.

Kirschner, P.A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. Learning and Instruction, 12, 1–10.

Kvale, S. (1997). Den kvalitativa forskningsintervjun (S.-E. Torhell, Trans.). Lund, Sweden: Studentlitteratur.

Lindgren, J., & Karlsson, M. (2001). WinSolGasWater (WinSGW) (version 2.2.4). Umeå, Sweden: Unit for Inorganic Chemistry, Department of Chemistry, Umeå University, Sweden.

Lundberg, B.K.S. (1998). Motivated learning—open experiments in chemistry. Progress in creativity and critical thinking. Stockholm: National Agency for Higher Education.

Lundberg, B.K.S., Berg, A., Geladi, P., & Johansson, E. (2002). Tertiary chemistry students' perception of learning and experiments. A principal component analysis of developmental positions. Unpublished manuscript, Umeå University, Sweden.

Mackenzie, A.M., Johnstone, A.H., & Brown, R.I.F. (2003). Learning from problem based learning. University Chemistry Education, 7, 1–14.

Moore, W.S. (1994). Student and faculty epistemology in the college classroom: The Perry schema of intellectual and ethical development. In K.W. Prichard & S.R. McLaran (Eds.), Handbook of college teaching: Theory and applications (pp. 46–67). Westport, CT: Greenwood Press.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. International Journal of Science Education, 25, 1049–1079.

Perry, W.G. (1970). Forms of intellectual and ethical development in the college years: A scheme. New York: Holt, Rinehart and Winston.

Perry, W.G. (1998). Forms of intellectual and ethical development in the college years: A scheme. San Francisco: Jossey-Bass.

Reid, N. (2003). Getting started in pedagogical research in the physical sciences. Hull, UK: Department of Chemistry, University of Hull.

Ryan, R.M., & Deci, E.L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. Contemporary Educational Psychology, 25, 54–67.

Skou, C.V. (2001). Qualitative media analyzer—software for the 21st century [Internet]. Skou, Sweden: Carl Verner.

Swaak, J., & de Jong, T. (1996). Measuring intuitive knowledge in science: The development of the What-If test. Studies in Educational Evaluation, 22, 341–362.

Swaak, J., Van Joolingen, W.R., & de Jong, T. (1998). Supporting simulation-based learning: The effects of model progression and assignments on definitional and intuitive knowledge. Learning and Instruction, 8, 235–252.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12, 257–285.

Sweller, J., van Merrienboer, J.J.G., & Paas, F.G.W.C. (1998). Cognitive architecture and instructional design. Educational Psychology Review, 10, 251–296.

Thomas, R., & Hooper, E. (1991). Simulations: An opportunity we are missing. Journal of Research on Computing in Education, 23, 497–514.

van Bruggen, J.M., Kirschner, P.A., & Jochems, W. (2002). External representation of argumentation in CSCL and the management of cognitive load. Learning and Instruction, 12, 121–138.

van Merrienboer, J.J.G., Kirschner, P.A., & Kester, L. (2003). Taking the load of a learners mind: Instructional design for complex learning. Educational Psychologist, 38, 5–13.

Veerman, A., Andriessen, J., & Kanselaar, G. (2002). Collaborative argumentation in academic education. Instructional Science, 30, 155–186.

Winberg, T.M. (in press). Assessing discourse quality in threaded discussion groups: A comparison of methods. Instructional Science.

Windschitl, M., & Andre, T. (1998). Using computer simulations to enhance conceptual change: The roles of constructivist instruction and student epistemological beliefs. Journal of research in science teaching, 35, 145–160.